TTS System for Coal Miners in MANET Based Disaster Management System

Abstract—Mobile Ad-hoc Networks (MANETs) are self-organized networks built up for moving objects. Logistic difficulties, economic issues, technology limitations make simulation the mean of choice in the validation of networking protocols for MANETs. It is quite interesting to bring down complexities of MANETs in realization of their characteristics both in simulation as well as real-time implementation by defining MANET for coal miners tracking applications within the geographic scenario specified. Wired-cum-wireless technology with mobility in network layer is used to achieve the desired goals of tracking and monitoring coal miners. The controlling station situated on the surface of the earth controls and monitors the activities of all the miners using the Text-To-Speech (TTS) system with different windows at the same time in their local language during disaster conditions. The motion of miners is monitored across a predefined geographical area and network capability of up to 15 miners. The focus is not only to track the miners, but also to establish an efficient disaster management system, in order to safeguard the lives of miners. The design of a prototype considering a base station and fifteen nodes (15 miners) is achieved. This paper gives a comprehensive introduction to state-of-the-art TTS synthesis application to disaster management system using sensor MANET.

Keywords: Mobile Ad-hoc Network, Access points, Nodes, Repeaters, Miners, TTS system, Formants, Grapheme, Phoneme

I. INTRODUCTION

A. Disaster Management System for Mines

The physical and geographical complexity of tunnels in underground mines impairs the visibility of environment and operating assets, including human beings. Irrespective of the acuteness of locating mineral though proper mining design and planning, the hazards of the sub-terrestrial world remains unpredictable. Most of the available technologies cannot predict an imminent disaster situation. Once the disaster strikes, it leaves very little window of opportunity to control the disaster from getting worse.

To overcome the difficulties for communication between miners, environmental monitoring, surveillance system etc. under mines, the paper explains an efficient disaster management system which monitors and tracks the miners at every instant and also helps in establishing an efficient wireless communication system between them, along with the most advantageous surveillance system [1].

B. TTS in Mine Disaster Management System

The controlling station sends commands to multiple miners at a time giving instructions in local language to miners which is easily understood as most of the miners are illiterates. The manpower required at the controlling station can also be reduced using this technology as this encourages the controlling station to send multiple messages at a time to different predetermined geographical areas underground. Text to Speech (TTS) synthesis, the artificial production of human speech and uses a speech synthesizer implemented in software or hardware, takes text as input and produces speech in a particular language.

II. RELATED WORK

A. Vehicular Ad-Hoc Network

VANET deals with moving nodes which are a tad quicker referred to vehicles (nodes). It works on the principle identification of a vehicle, which is assigned as Mobile-IP address to a vehicle. The system tracks the vehicle through access points, which are established at various locations in lane or parks or in large campus. The carrier is a navigation server that connects with multiple clients (nodes) and is also responsible for the clients request data. The client control panel system running on client side helps the user for identifying current location, destination location, landmark and the distance to be traveled. Traffic intensity of each lane is notified through carrier access point. Shivmurthy, et al, explains such a situation wherein the tracking of vehicles is done using two simulators, Network Simulator (NS) and Simulation Urban Mobility (SUMO).

B. Mobile Ad-Hoc Network for Coal Miners Disaster Management

1) Simulation of Network Scenario: Routing information in NS is generated based on the connectivity of the topology where nodes are connected to one another through Links. Base station here is the controlling station situated on the earth surface and it is wired to the access points situated under mines at some predefined static places. The mining area is completely installed with the MANETs environment, each miner acting as a mobile node and access point itself being a static node. Base Station Node is created which plays the role of a gateway for the wired and wireless domains.
The Base Station Node is essentially a hybrid between a Hierarchical node (Hier Node) and a Mobile Node. The base station node is responsible for delivering packets in and out of the wireless domain. Each wireless domain along with its base-station would have a unique domain address assigned to them. All packets destined to a wireless node would reach the base-station attached to the domain of that wireless node, who would eventually hand the packet over to the destination (mobile node). And mobile nodes route packets, destined to outside their (wireless) domain, to their base-station node. The mobile nodes in wired-cum-wireless scenario are required to support hierarchical addressing/routing. The Mobile Node looks exactly like the Base Station Node.

2) Hardware Implementation of Sensor Systems: The mobile unit which is fixed in the headgear of the miner acts as the node of the MANET. The mobile unit designed is as shown in the figure. The mobile unit will have different types of sensors to gather the information of environmental conditions and send to the base unit which is fixed above the earth surface. Microcontroller to process the information, Analog to Digital Converter (ADC) to convert the analog electrical signal outputs of the sensors into digital signals for the proceeding operations. The mobile system also will have the communication equipments and multimedia devices. The transceiver helps in the communication process between miners and between miner and the base unit operator at the controlling station.

3) Hardware Implementation of Surveillance System: Recorder: The circuit diagram shown (Figure 3) sums up the design and implementation methods we have followed for bringing the surveillance system using PIR sensor and coupling it with the storage media. The circuit develops around the AT Mega 32, and most interfacing resolves to straight connections, as all AVR I/O pins provide programmable pull-ups and respectable output current. One notable exception is the PIR movement sensor which needs an independent power regulator. The camera module connects to the UART RX and TX pins. Communication runs at 115200 baud, thus a 7.3728 MHz crystal for exact bit timing is used and the SD-CARD connects directly to the SPI port pins.

The connector features extra signals indicating card insertion status and the position of the write-protect tab of the card. Then route them to two spare I/O pins. A different kind of switch, the micro switch detecting when the lid is opened, and the external trigger switch connect in the same way to the AVR. However, we select one of the three AVR interrupt pins for the external trigger, in order to capture and flag automatically any transition which applies to the PIR input also. The latter pin comes from infrared receiver IC. Four LEDs showing device status take half of the port A. A spare bit on port C drives the relay output by means of the typical transistor & diode network.

The snapshots of the simulated environment comprising of nodes (miners), access points and base station are shown (Figure 2).

The small circles indicate the nodes of the network environ-
ment, the access point and base station are shown in hexagons. The circles around the nodes explain the coverage areas of the individual nodes (Figure 5). The nodes are designed to be in continuous motion, communicating with the neighbors. Each node transfers the information about its positions (to the base station every instant (Figure 6 & 7). The miners are identified at the base station by the unique IP address. The controlling station gives out the warning signals as well as other required information through the access point. The access points act as repeaters, increasing the signal strength in the mining environment. The snapshots are taken from the Network animator. Each and every node here is being tracked and the packet exchange between miners and the base station are shown (Figure 8). The node 4, 11, and 10 perfectly explains the packet exchange, acknowledgements exchange and how the safety of miners is achieved (Figure 9). MANET simulations were carried out in a topological area of 100mx100m. Under this area include 15 to 50 miners represented as nodes and each has a wireless IP address with an ability pattern randomly designed. The data exchange between the nodes and the access points is shown in simulation results. This area simulated under a base station and various access points. Mobile Host that moves to other domain communicates through Mobile IP. The mobility of Mobile Hosts is varied in the range of 0m/s to 1m/s.

The work of installing sensors initially with pressure, gas and temperature sensors along with camera in the headgear of the miners was successful. The Miners Tracking in a defined geographical area was logged into a separate information file for 15 miners. In the mobility observation shown are the screen shots of the nodes (miners) in motion as they take random path and their mobility can be seen. Here each node is represented by a circle, access points and base station by a hexagon. The big circles around nodes indicate their radio range by default 100m is used for each node. The path taken can be analyzed using wireless trace file.

### III. RESULTS OF HARDWARE SYSTEMS

The prototype designed has one mobile unit and a base unit. Every node is given its own IP address and the base station keeps on receiving the signals from all the nodes. Getting the information (Figure 10), it updates the log file about the location of the miner, the surrounding conditions of the particular miner is displayed. Based on the conditions received, the base unit decides which miner has to be warned and to that miner having the particular IP address, from which dangerous conditions are being monitored, the warning signals are sent. If the miner to which the warnings has to be sent is out of the network coverage area, the network sends the signals to the destined miner via the neighboring miner as he still in the network area of the lost miner. The controller can also receive and send the voice signals from the miners and also some external warning signals if needed. The designed hardware unit efficiently transmits and receives to a range of 100 mts. The conditions considered are above the earth surface, normal environmental conditions as initial work. The sensors installed in the headgear uses microcontroller for interrupts and depending on the microcontroller used, the number of sensors can be increased thereby increasing the monitoring efficiency. As the network protocol used being DSDV the loss of data is minimal and efficient communications links are established.

The power loss is made much minimum by making the communication wired till the access points from the base station. The photo (Figure 11) above shows the ceiling view of
Fig. 9. Ceiling view of a room captured by the Self Recording Surveillance System Using PIR Sensor.

one of the living-room captured by the SELF RECORDING SURVEILLANCE SYSTEM USING PIR SENSOR. Mileage can vary, but should be able to get about 50,000 frames at 320x200 (like the one below), or 25,000 at 640x480, using a 1 GB card which corresponds to more than 40 hours of overall recording. In our case, just 20 minutes on average giving an impressive 120 days of storage capacity. The SELF RECORDING SURVEILLANCE SYSTEM USING PIR SENSOR fulfills all our needs. From experience it would be presumptuous to consider it a finished product. Extensive testing, like verification against the most commonly available cards and PC readers, would be a huge task. Nonetheless, its potential should be easy to appreciate, even in this homemade incarnation. The SELF RECORDING SURVEILLANCE SYSTEM USING PIR SENSOR merges two mainstream technologies flash memory and image sensors to create something new and practical. Replacing the LCD with a voice interface is both cost-effective and optimal from a ease-of-use perspective. The choice of the Mega32 processor is a perfect fit. It handles easily its workload, and almost all its hardware features are used. Also the sizes of FLASH and RAM are just right for the application.

IV. DESIGN AND IMPLEMENTATION OF TTS SYSTEM

A. Design of TTS System

1) Speech Segmentation: Speech segmentation is the process of identifying the boundaries between words, syllables, or phonemes in spoken natural languages. Speech segmentation is an important sub problem of speech recognition, and cannot be adequately solved in isolation. As in most natural language processing problems, one must take into account context, grammar, and semantics, and even so the result is often a probabilistic division rather than a categorical.

2) Grapheme to Phoneme: In a method for grapheme to phoneme conversion of a word which is not contained as a whole in a pronunciation lexicon, the word is firstly decomposed into sub words. The sub words are transcribed and chained. As a result, interfaces are formed between the transcriptions of the sub words. The phonemes at the interfaces must be changed frequently. Consequently, they are subjected to recalculation.

3) Database: A speech corpus is a database of speech audio files and text transcriptions in a format that can be used to create acoustic models (which can then be used with a speech recognition engine). There are two types of Speech Corpora: One the Read Speech which includes book excerpts, broadcast news, lists of words, sequences of numbers and the other spontaneous speech which includes dialogues between two or more people, narratives, map-tasks, appointment-tasks - two people try to find a common meeting time based on individual schedules.

4) Concatenation: In a speech synthesizer apparatus, a weighting coefficient training controller calculates acoustic distances between one target phoneme from the same phoneme and the phoneme candidates other than the target phoneme based on first acoustic feature parameters and prosodic feature parameters. Then, a speech unit selector searches for a combination of phoneme candidates which correspond to a phoneme sequence of an input sentence and which minimizes a cost including a target cost representing approximate costs between a target phoneme and the phoneme candidates and a concatenation cost representing approximate costs between two phoneme candidates to be adjacently.

B. Implementation

The database used consists of phonetically rich words and sentences completely available in the database and these are segmented into units. Kannada and Tamil are two local(regional) spoken languages among many Indian languages which has its own script.

Every segment might have more than one choice in the database. The challenge would be to select the best choice for concatenation so as to get a sound form closer to natural sound. The three methods for unit selection are:

1) Random Selection
2) Pitch based selection
3) Formant based selection

1) Random selection: In this method among the many choices for a particular segment in the database, a segment is picked randomly for concatenation.

2) Pitch Based Selection: The time duration of one glottal cycle is called pitch period. The reciprocal of pitch period is called pitch. In conversational speech, during vowel sounds, typically one to four pitch periods can be seen. Since words are segmented at the middle of the vowel, units can be selected based on the pitch of the units.
3) **Formant based Selection:** The spectral peaks of the sound spectrum are called formants. They correspond to the resonant frequencies in the vocal tract. The vocal tract transfer function can be expressed either in product or partial fraction expansion form:

\[ H(z) = \frac{A}{\prod (1 - c_k z^{-1})(1 - c_k^* z^{-1})} \]  

(1)

where the roots are complex conjugate poles inside the unit circle. More specifically, when vocal tract is modeled as a time-invariant all-pole linear system then, a pole at \( z_0 = r_0 e^{j\omega_0} \) corresponds approximately to a vocal tract formant. **Formant based segment selection** involves comparing the formants of the vowels of the segment to be concatenated. Comparison is done for first formants (F1) and second formants (F2). This involves extraction of first formants and second formants.

The flowchart for the above mentioned procedure is as shown(Figure 13) :

![Flow Chart](image)

**C. Methods for Formant Extraction**

Spectral peak picking method and Root extraction method. A brief description of different methods of formant extraction is given below:

1) **Root Extraction Method:** In this method, like the spectral peak picking method, we first compute linear prediction (LP) coefficients and obtain the prediction-error filter \( A(z) \). Comparing with (1), we can easily find that the roots of this polynomial \( A(z) \) correspond to the poles of the vocal tract system. Thus, we can obtain candidates for formants by solving \( A(z) = 0 \), using numerical methods. When poles are kept sufficiently apart, and one of these poles, \( z = r_0 e^{j\omega_0} \), forms a formant, the formant frequency \( F \), and the formant bandwidth \( B \) can be represented by the equation[1]:

\[ F = \frac{f_s}{2\pi} \frac{\phi_0}{\pi} \]  

(2)

\[ B = -\frac{f_s}{\pi} \ln(r_0) \]  

(3)

where \( r_0 \) is the magnitude of the pole, \( \phi_0 \) is the phase of the pole, \( f_s \) is the sampling frequency, \( F \) is the formant frequency, and \( B \) is the 3-dB formant bandwidth. Thus, if we find the roots of the prediction-error polynomial, we can obtain the formant frequencies using (1). In addition, we can get the bandwidth information from (2).

![Location of poles in speech signal](image)

2) **Implementation:** Steps followed for extraction of formants using root extraction method are as follows: Input speech signal is low pass filtered with a cut-off frequency of 3 kHz in order to reduce the effects of higher formants. Linear Prediction Coefficients are calculated for the given speech signal. Linear Prediction Coefficients are the coefficients of the polynomial that forms the vocal tract transfer function. Solving for the roots of the polynomial to get formant frequency.

3) **Spectral Peak picking method:** The spectral peak picking method and its variants have been widely used for formant extraction. In most cases, instead of the short-term spectrum itself, smoothed spectra, such as linear prediction (LP) spectrum or spectrally smoothed spectrum are often employed. However, LP spectra are more often used for this purpose, since they show conspicuous peaks. Additionally, it has been verified that the prediction-error polynomial obtained from LP coefficients is closely related to the vocal tract filter, which generates the formants. Figure 13 shows the short-term spectrum of the “ae” sound, and Figure illustrates the LP spectrum of this signal.
D. Method to compute LP Spectrum:

Vocal tract system \( H(z) \) can be approximately modeled as follows:

\[
H(z) = \frac{G}{\sum_{k=0}^{N_{lp}} a_k z^{-k}}
\]  

(4)

Let us denote LP coefficients of a short-term speech signal by \( a_k \), \( 0 \leq k \leq N_{lp} \), where \( N_{lp} \) is the prediction order. From these LP coefficients, we can construct the following prediction-error filter:

\[
A(z) = \sum_{k=0}^{N_{lp}} a_k z^{-k}
\]  

(5)

As mentioned above, previous studies show that the vocal tract filter is modeled as an all-pole system, and the vocal tract filter in (1) can be obtained from the prediction-error filter in (2) which is also known as the inverse filter (IF). By performing FFT of sufficient order like 256 or 512, on the zero-padded LP coefficients, we can obtain a reasonable amplitude spectrum of the vocal tract system shown in (1). The spectrum obtained by the above-mentioned procedure, LP spectrum. As the name suggests, this type of formant extractors tries to find resonances on the spectrum.

Advantages of spectral peak picking: They show relatively reliable results. They do not require much computation.

1) Implementation of the method: The flowchart for spectral peak picking method is as shown (Figure 16):

The algorithm is as explained: Speech segment is normalized to get the discrete values of the speech in the range ±1. LP coefficients are calculated using Levinson’s and Durbin’s algorithm which uses autocorrelation method to calculate linear prediction coefficients. Since the first formant frequency is always within the range of 50 to 1600 Hz, the range of frequencies over which computation is made is fixed. The angular frequency \( \omega = 2\pi f \) is calculated for every frequency considered. The amplitude is calculated by computing the magnitude of

\[
A = \frac{1}{a_k z^{-k}}
\]  

(6)

where \( z = e^{j\omega} \).

For the computation of magnitude, values of \( \sin(\omega a_k) \) and \( \cos(\omega a_k) \) is needed. Once the amplitude of the spectrum is obtained, maximum value and the corresponding frequency is
TABLE I
FORMANT EXTRACTION RESULTS FOR A SPEECH SEGMENT (SPEECH FILE: TEST1.WAV) USING MATLAB.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Spectral peak picking</th>
<th>Root Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>F1</td>
<td>F2</td>
</tr>
<tr>
<td>0.52</td>
<td>609.08</td>
<td>1179.58</td>
</tr>
<tr>
<td>0.53</td>
<td>614.8</td>
<td>1154.5</td>
</tr>
<tr>
<td>0.54</td>
<td>587.6</td>
<td>1177.6</td>
</tr>
<tr>
<td>0.55</td>
<td>613.1</td>
<td>1151.6</td>
</tr>
<tr>
<td>0.56</td>
<td>634.0</td>
<td>1079.1</td>
</tr>
<tr>
<td>0.57</td>
<td>602.6</td>
<td>1052.7</td>
</tr>
<tr>
<td>0.58</td>
<td>562.3</td>
<td>1087.4</td>
</tr>
<tr>
<td>0.59</td>
<td>517.7</td>
<td>937.8</td>
</tr>
<tr>
<td>0.60</td>
<td>527.5</td>
<td>1082.3</td>
</tr>
</tbody>
</table>

TABLE II
FORMANT EXTRACTION RESULTS FOR A SPEECH SEGMENT (SPEECH FILE: TEST1.WAV) USING C AND MATLAB.

<table>
<thead>
<tr>
<th>Time</th>
<th>Matlab</th>
<th>C Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.52</td>
<td>609.08</td>
<td>609.0</td>
</tr>
<tr>
<td>0.53</td>
<td>614.8</td>
<td>614.7</td>
</tr>
<tr>
<td>0.54</td>
<td>587.6</td>
<td>587.6</td>
</tr>
<tr>
<td>0.55</td>
<td>631.5</td>
<td>631.5</td>
</tr>
<tr>
<td>0.56</td>
<td>634.0</td>
<td>634.0</td>
</tr>
<tr>
<td>0.57</td>
<td>562.3</td>
<td>562.3</td>
</tr>
<tr>
<td>0.58</td>
<td>517.7</td>
<td>517.7</td>
</tr>
<tr>
<td>0.59</td>
<td>527.5</td>
<td>527.5</td>
</tr>
</tbody>
</table>

computed. The procedure is repeated to get the precise value of formant frequency.

2) Unit selection in TTS using extracted formants: We are using F1 and F2 values as a parameter for continuity metric along with the pitch metric. A combination of the pitch and formant will be used to obtain an optimal cost for continuity metric. The code for spectral peak picking is integrated in the source code of TTS system.

V. RESULTS OF TTS SYSTEM

Formant frequencies are extracted successfully using the methods mentioned and the results are satisfactory. It is inferred that root extraction method involves cumbersome computations and is time consuming. Spectral peak picking method gives reliable results and is computationally simple. The methods are implemented in MATLAB. The spectral peak picking method is implemented in C (eclipse C/C++ platform).

The pitch track function which computes pitch was replaced by formant track function. The speech wave file is synthesized successfully. The units selected are smooth and the prosody is better compared to using pitch track function. The time consumed is more than the pitch track but negligible. Peak picking method results in peak merger problem in some cases.

Formant extraction results for a speech segment (speech file: test1.wav) using Matlab are tabulated as shown in Table 1.

Formant extraction results for a speech segment (speech file: test1.wav) using C and Matlab comparison is as shown in Table 2.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we try to bring out the good use TTS System in the field of mine disaster management control using Sensor Mobile Ad-Hoc Networks. The system not only guides the miners using disaster conditions tracks of a particular node but also keeps its trace at all instants of time.

We intend to incorporate the solution for peak merger problem involved in extraction of formants. Our work also intends to achieve the system working even during power failure for very long duration and bring out the overall model in the form of chip, which eases problems of bulkiness and large area.

REFERENCES